Meta-analysis of published associations versus pooled analysis by large consortia

ACJW Janssens, AM González-Zuloeta Ladd, S López-León, JPA Ioannidis, BA Oostra, MJ Khoury, CM van Duijn

Erasmus University Medical Center Rotterdam, Rotterdam, the Netherlands
University of Ioannina School of Medicine, Ioannina , Greece
Tufts University School of Medicine, Boston, USA
Centers for Disease Control, National Office of Public Health Genomics, Atlanta, USA

Gene-disease association studies

Genetic basis of complex diseases

- Multiple genes involved
- Each gene very minor effect

To identify susceptibility genes with minor effects:

- Many replication studies → meta-analyses
- Studies with very large sample size → consortia

Erasmus MC

Consortia on gene-disease associations

Name	N cases	(Total)
Breast Cancer Association Consortium	<30,000	(<60,000)
Genetic Markers for Osteoporosis consortium		(<26,000)
International Lymphoma Epidemiology Consortium	<10,200	(<18,000)
Consortium of Investigators of Modifiers of BRCA1/2	15,000	carriers
Breast and Prostate Cancer Cohort Consortium	5,600	breast
	6,700	prostate
International Consortium for Prostate Cancer Genetics	1,250	families
DiaGen Consortium (Type 2 diabetes)	<2,600	(<5,400)
Consortium on Genetics of Schizophrenia		(recruiting)
Type 1 Diabetes Genetics Consortium		(recruiting)

Updated: july 2007

Erasmus MC

Consortia on gene-disease associations

Advantages consortium approach

- Larger sample size
- Access to unpublished data
- Harmonization of criteria and definitions, and standardization of genotype technology → reduce between-study heterogeneity

Disadvantage:

- Lot of work, compared to meta-analyses
- Not all research groups are involved

Erasmus MC

Research question

Do consortium analyses and meta-analyses of published data yield same results?

Strategy:

- Choose publication of consortium with gene-disease associations
- Perform literature search on same polymorphisms
- Perform meta-analyses
- Compare population size, between-study heterogeneity, potential sources of bias, results

Erasmus MC

Commonly Studied Single-Nucleotide Polymorphisms and Breast Cancer: Results From the Breast Cancer Association Consortium

The Breast Cancer Association Consortium

Journal of the National Cancer Institute, Vol. 98, No. 19, October 4, 2006

A common coding variant in *CASP8* is associated with breast cancer risk

Angela Cox[1,33], Alison M Dunning[2,33], Montserrat Garcia-Closas[3,33], Sabapathy Balasubramanian[1], Malcolm W R Reed[1], Karen A Pooley[2], Serena Scollen[2], Caroline Baynes[2], Bruce A J Ponder[2], Stephen Chanock[3], Jolanta Lissowska[4], Louise Brinton[3], Beata Peplonska[5], Melissa C Southey[6], John L Hopper[6], Margaret R E McCredie[7], Graham G Giles[8], Olivia Fletcher[9], Nichola Johnson[9], Isabel dos Santos Silva[9], Lorna Gibson[9], Stig E Bojesen[10], Børge G Nordestgaard[10], Christen K Axelsson[10], Diana Torres[11], Ute Hamann[11], Christina Justenhoven[12], Hiltrud Brauch[12], Jenny Chang-Claude[13], Silke Kropp[13], Angela Risch[13], Shan Wang-Gohrke[14], Peter Schürmann[15], Natalia Bogdanova[16], Thilo Dörk[15], Rainer Fagerholm[17], Kirsimari Aaltonen[17,18], Carl Blomqvist[18], Heli Nevanlinna[17], Sheila Seal[19], Anthony Renwick[19], Michael R Stratton[19], Nazneen Rahman[19], Suleeporn Sangrajrang[20], David Hughes[21], Fabrice Odefrey[21], Paul Brennan[21], Amanda B Spurdle[22], Georgia Chenevix-Trench[22], The Kathleen Cunningham Foundation Consortium for Research into Familial Breast Cancer, Jonathan Beesley[22], Arto Mannermaa[23], Jaana Hartikainen[23], Vesa Kataja[23], Veli-Matti Kosma[23], ...negien Broeks[25], Marjanka K Schmidt[25], Frans B L Hogervorst[25], ...[27], Dong-Young Noh[26], Sei-Hyun Ahn[28], Sara Wedrén[29], ...[31], Gloria Ribas[31], Anna Gonzalez-Neira[31], Javier Benitez[31], ...nder[32], Jeffery P Struewing[32], Paul D P Pharoah[2] & ...ciation Consortium

VOLUME 39 | NUMBER 3 | MARCH 2007 NATURE GENETICS

BCAC:

- 20 research groups
- Individual-level data up to 30,000 patients
- Case-control studies

Erasmus MC

Breast Cancer Association Consortium

Initial analyses: available data up to 16,000 patients

16 genetic polymorphisms: No association: n=12

Association at $p < 0.10$: n=4

Additional genotyping in remaining groups:

Available data up to 30,000 patients

4 polymorphisms: No association: n=2

Association at $p < 0.05$: n=2

CASP8 (and *TGFB1*)

Erasmus MC

Meta-analyses of published studies

Databases: PubMed, HuGENet, Web of Science

Search strategy: 'breast cancer' AND <name of gene>

Inclusion:

- Female breast cancer patients
- Controls from the general population
- Case-control design
- Reported in English

Exclusion:

- Data were re-used on the same polymorphism
- Control genotype distributions not in HWE
- Incomplete reporting of genotype frequencies

Erasmus MC

Results

Included:	115	publications
Excluded:	5	incomplete reporting
	1	tumor DNA
Remains:	109	publications
Included	168	datasets on the 16 polymorphisms
Excluded:	4	re-used in larger study
	2	gene not polymorphic
	11	controls distributions not in HWE
Available:	151	sets of data

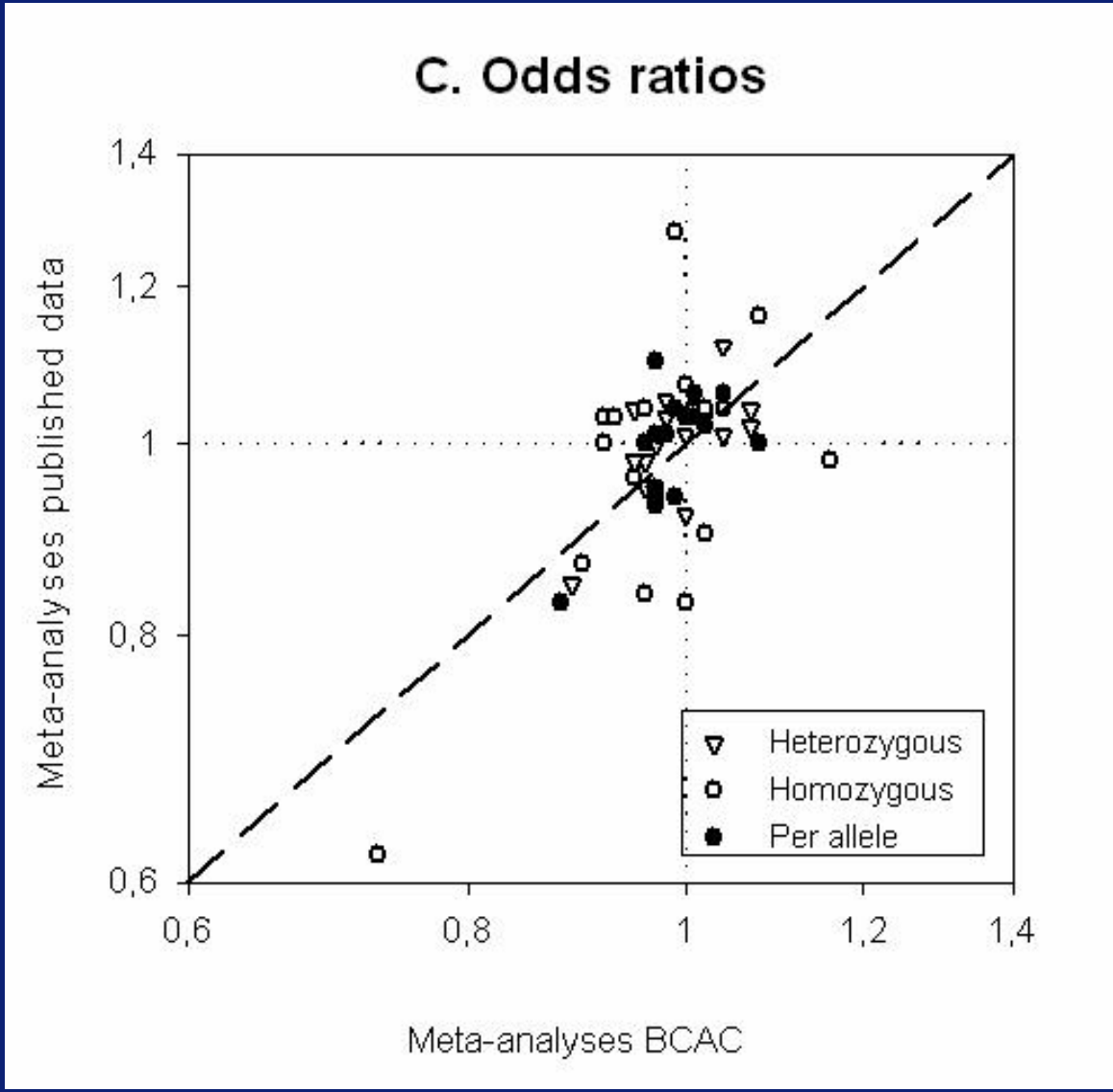
Erasmus MC

Results meta-analyses published data

Gene	Alteration	Number of Studies	Controls	Cases	Heterozygotes		Homozygotes		Per allele	
					OR (95% CI)	I2	OR (95% CI)	I2	OR (95% CI)	I2
ADHC1	I350V	4	2390	2130	1.00 (0.87-1.15)	14	**0.83 (0.69-1.00)**	0	0.94 (0.86-1.02)	0
AURKA	F31I	9	9011	7294	1.04 (0.96-1.13)	0	**1.28 (1.06-1.54)**	46*	**1.10 (1.01-1.19)**	58*
BRCA2	N372H	8	14387	14065	1.04 (0.99-1.09)	0	1.04 (0.93-1.16)	19	1.03 (0.99-1.06)	0
CASP8	D302H	3	3591	3288	**0.85 (0.76-0.95)**	0	**0.62 (0.42-0.89)**	0	**0.83 (0.75-0.92)**	0
ERCC2	D312N	12	7821	9414	0.98 (0.86-1.11)	63**	0.84 (0.68-1.05)	70***	0.94 (0.84-1.05)	77***
IGFBP3	C(-202)A	9	12294	9774	1.01 (0.94-1.09)	13	1.00 (0.90-1.11)	32	1.01 (0.96-1.06)	30
LIG4	D501D T/C	3	4113	4520	0.95 (0.87-1.05)	0	0.90 (0.59-1.36)	59*	0.95 (0.87-1.04)	9
PGR	V660L	9	11646	10652	1.04 (0.95-1.13)	33	1.03 (0.73-1.46)	56*	1.02 (0.92-1.13)	60**
SOD2	V16A	12	11141	9991	1.03 (0.95-1.10)	11	1.04 (0.92-1.17)	40*	1.01 (0.96-1.07)	28
TGFB1	L10P	17	16308	9331	1.02 (0.93-1.12)	41*	0.98 (0.86-1.12)	46*	1.00 (0.94-1.06)	46*
TP53	R72P	14	8218	7569	1.03 (0.91-1.17)	61**	1.04 (0.87-1.25)	48*	1.03 (0.94-1.13)	63***
XRCC1	R399Q	21	14479	13320	1.05 (0.98-1.12)	21	1.07 (0.96-1.19)	32*	1.04 (0.99-1.10)	47*
XRCC2	R188H	7	9723	10427	0.98 (0.89-1.08)	24	1.03 (0.62-1.70)	40	1.00 (0.89-1.11)	47*
XRCC3	5'UTR A/G	4	6563	6303	**1.12 (1.00-1.24)**	48	0.96 (0.81-1.15)	0	1.06 (0.98-1.14)	26
XRCC3	IVS5-14	4	6682	6270	0.92 (0.82-1.03)	58*	**0.87 (0.75-1.00)**	33	0.93 (0.85-1.02)	64*
XRCC3	T241M	15	13370	14255	1.01 (0.95-1.07)	7	**1.16 (1.04-1.28)**	30	**1.06 (1.01-1.12)**	39*

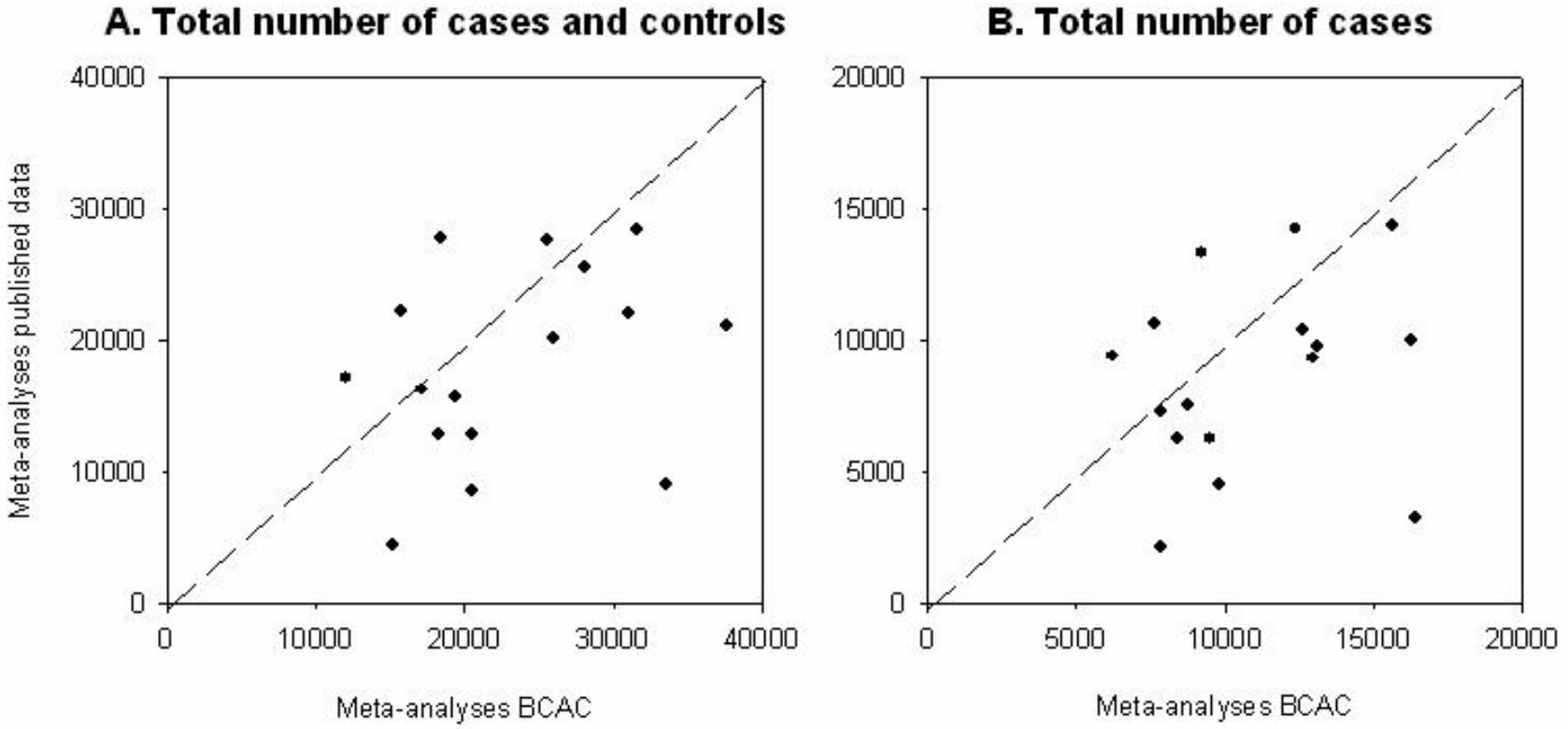
Erasmus MC

Differences in results



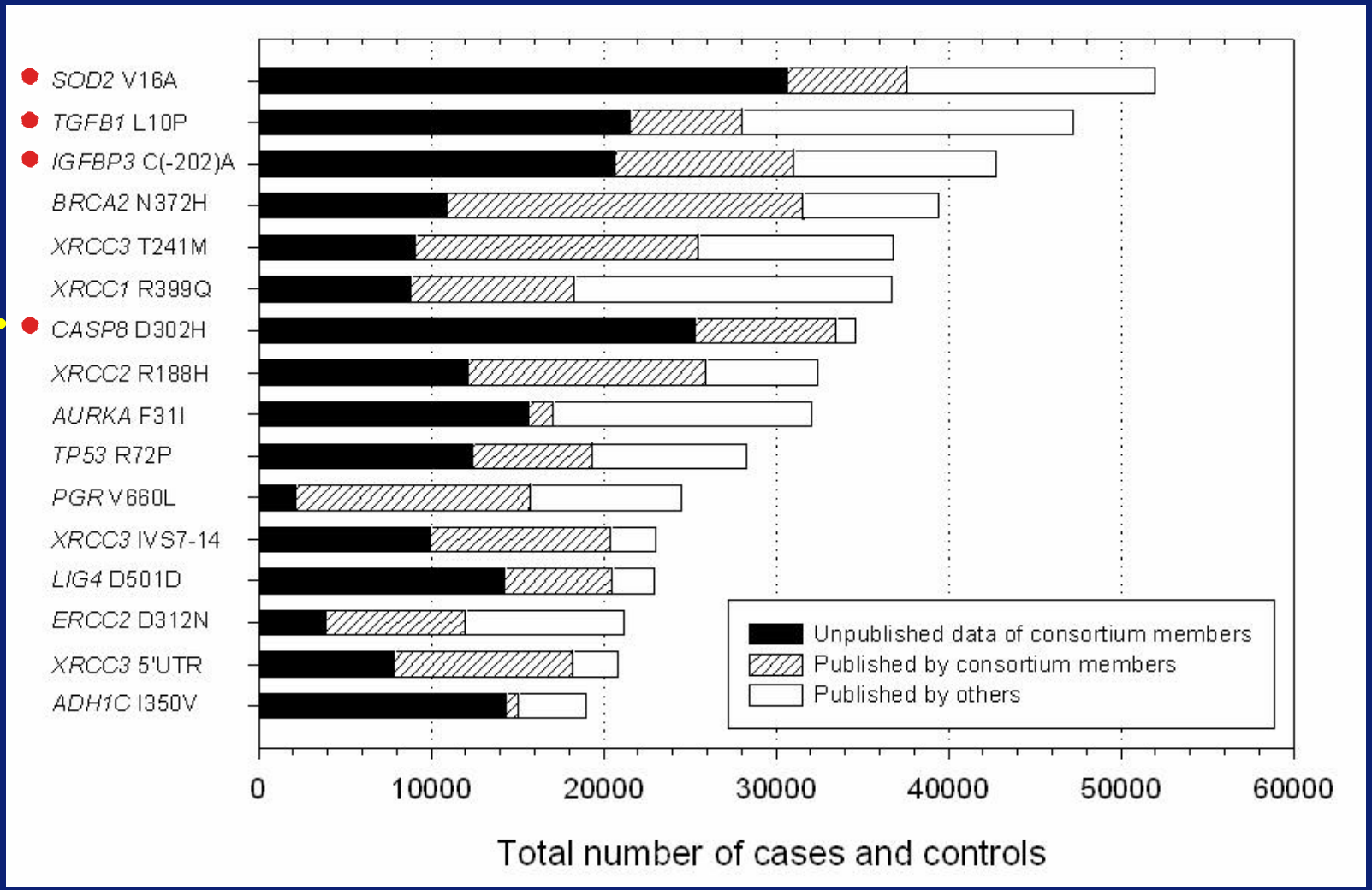
Erasmus MC

Differences in population size



Erasmus MC

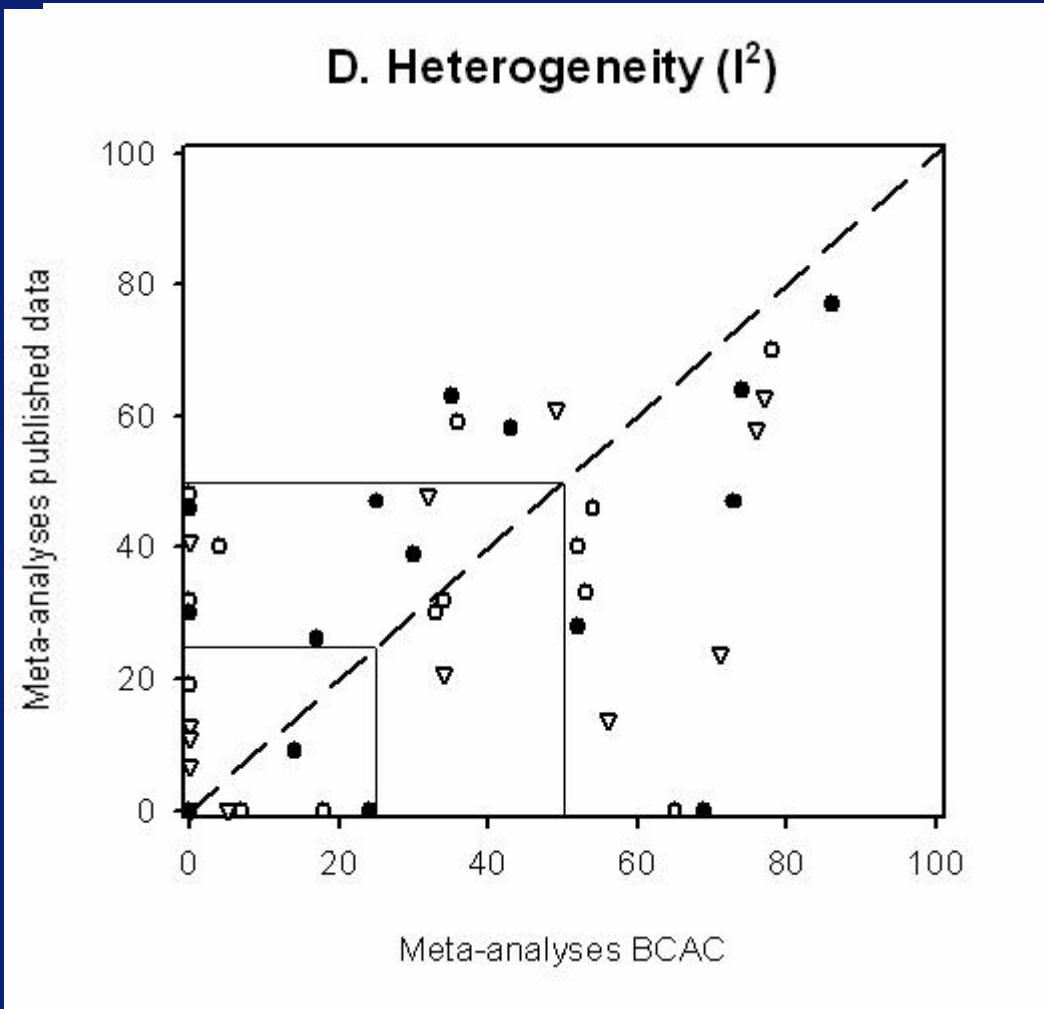
Differences in populations included



- Additional genotyping by consortium members

Erasmus MC

Differences in heterogeneity



Erasmus MC

Grading epidemiological evidence of association

I Amount of evidence

II Replication consistency

III Protection from bias

Assessment of cumulatiive evidence on genetic associations. Interim guidelines. Ioannidis, […], Khoury. Am J Epidemiol. 2007 (in press)

Each graded A, B, C

AAA	Strong epidemiological evidence of association
B** (no C)	Moderate evidence
C**	Weak evidence

Erasmus MC

Grading epidemiological evidence of association

I Amount of evidence: number of cases + controls in smallest genotype category:

A: > 1,000

B: 100-1,000

C: <100

Results:

- No differences between BCAC and MA-publ
- All heterozygous and per allele analyses: A
- 10/16 homozygous analyses: A, rest B

Erasmus MC

Grading epidemiological evidence of association

II Replication consistency, basically:

A: I^2 <25

B: I^2 25-50

C: I^2 >50 or Non statistically significant association

Results:

- Most C, because of No association
- MA-publ: 39/48 analyses: C
- BCAC: 40/48 analyses C

Erasmus MC

Grading epidemiological evidence of association

III Protection from bias:

A: Bias could affect magnitude, but not presence of association

B: No obvious bias, but insufficient information for A

C: Clear presence of bias that can affect even presence of association

Exceptions:

Consortium assumed grade A

Meta-analyses with OR>1.15 assumed grade A

Bias was only investigated for polymorphism that did not receive grade C for amount of evidence and replication consistency

Results: MA-Publ: 4/9 analyses A, rest C

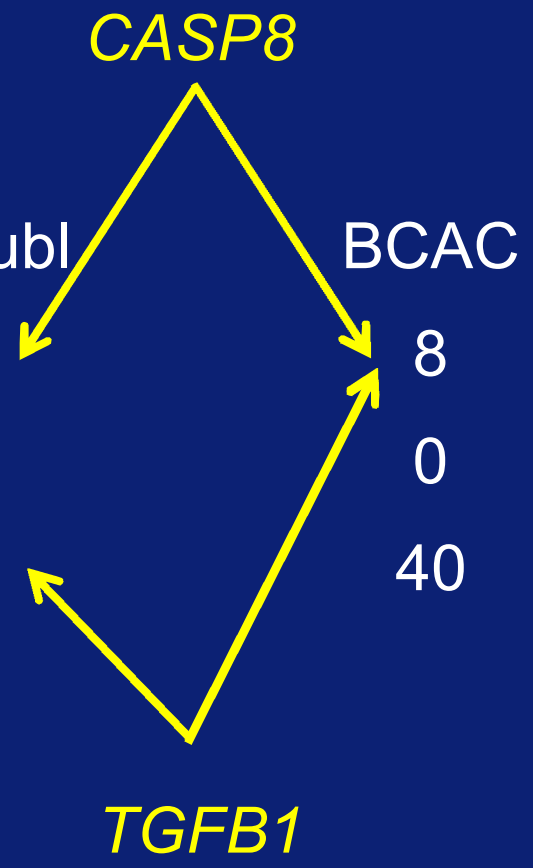
Erasmus MC

Grading epidemiological evidence of association

Final results:

Out of 48 analyses

		MA-publ	BCAC
AAA	Strong	4	8
B** (no C)	Moderate	0	0
C**	Weak	44	40



Erasmus MC

Summary of results

- Both approaches identified *CASP8*, both graded with strong evidence for association
- Consortium, but not meta-analyses of published data, showed moderated association for *TGFB1*

- When all data combined: *CASP8* associated, but not *TGFB1*

Erasmus MC

Consortia on gene-disease associations

Two types:

- Prospective: agree on definitions and criteria prior to data collection
  best, but costly and time-consuming

- Retrospective: combining available data
  most common

Erasmus MC

Retrospective consortia

Combine available case-control or cohort data of consortium members

- Access to unpublished data

But:

- No (or limited) harmonization of inclusion-exclusion criteria
- No (or limited) harmonization of diagnostic criteria
- No standardization of genotype technology
- Not all research groups participate

Erasmus MC

Differences between BCAC studies

- Postmenopausal versus premenopausal (age range e.g. < 50 or 44-91)
- Unilateral versus bilateral breast cancer
- Familial cases versus sporadic
- Screened control populations versus unscreened
- Hospital-based controls versus population-based
- Genotype platforms (Taqman, Illumina, enzyme-based assays)

Conclusion

- Meta-analyses of published data identified same genetic variants as consortium analyses
- Meta-analyses of published data and consortium analyses may provide complementary insights, despite the methodological issues concerning published data meta-analyses
- Further comparisons are needed to demonstrate the generalizability of this conclusion, both for retrospective and prospective consortia

Erasmus MC